# Cyclistic Bike-share Case Analysis

Lora Yang

2022-04-02

## Introduction

**In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.**

**Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.**

**Marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends.**

## process 1:Ask

## What's the problem?

***The marketing analyst needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics.***

## Business Task

***Design marketing strategies aimed at converting casual riders into annual members by finding out the the differences between casual and annual riders.***

## Key Stakeholders

***Cyclistic executive team***

***Lily Moreno, The director of marketing and my manager***

***Cyclistic marketing analytics team***

## Process 2: Prepare

## Data download

## Data organization

**3** ***csv.files***

## Source of Data

***Data was collected from Motivate International Inc under this license https://ride.divvybikes.com/data-license-agreement***

## Sort, filter and clean the Data use spreedsheet

***trim the spaces, delete the columns not related with the analysis***

## library the required packages

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(dplyr)
library(ggplot2)
library(readr)
```

### loading data from 01/2021 to 03/2021

```
m1_2021 <- read_csv("/cloud/project/cyc/202101-divvy-tripdata.csv")
```

```
## Rows: 96834 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): rideable_type, started_at, ended_at, start_station_name, member_casual
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
m2_2021 <- read_csv("/cloud/project/cyc/202102-divvy-tripdata.csv")
```

```
## Rows: 49622 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): rideable_type, started_at, ended_at, start_station_name, member_casual
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
m3_2021 <- read_csv("/cloud/project/cyc/202103-divvy-tripdata.csv")
```

```
## Rows: 228496 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): rideable_type, started_at, ended_at, start_station_name, member_casual
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

### Check the column names for the consistency

***str every format, display result***

### combine all data to one data frame

```
all_trip_share<-bind_rows(m1_2021, m2_2021, m3_2021)
```

## Process 3: Process

### clean data and add data

***head(all_trip_share) and drop all NA***

```
Cleaned_all_bike_share<-drop_na(all_trip_share)
```

***Add new variables month, day, year, day_of _week***

```
Cleaned_all_bike_share$month <- format(as.Date(Cleaned_all_bike_share$started_at),"%m")
Cleaned_all_bike_share$day <- format(as.Date(Cleaned_all_bike_share$started_at), "%d")
Cleaned_all_bike_share$year <- format(as.Date(Cleaned_all_bike_share$started_at), "%Y")
Cleaned_all_bike_share$day_of_week <- format(as.Date(Cleaned_all_bike_share$started_at), "%A")
```

## Process 4: analysis

### *Add new column length of ride*

```
Cleaned_all_bike_share$Length_of_ride<-difftime(Cleaned_all_bike_share$ended_at, Cleaned_all_bike_share$
```

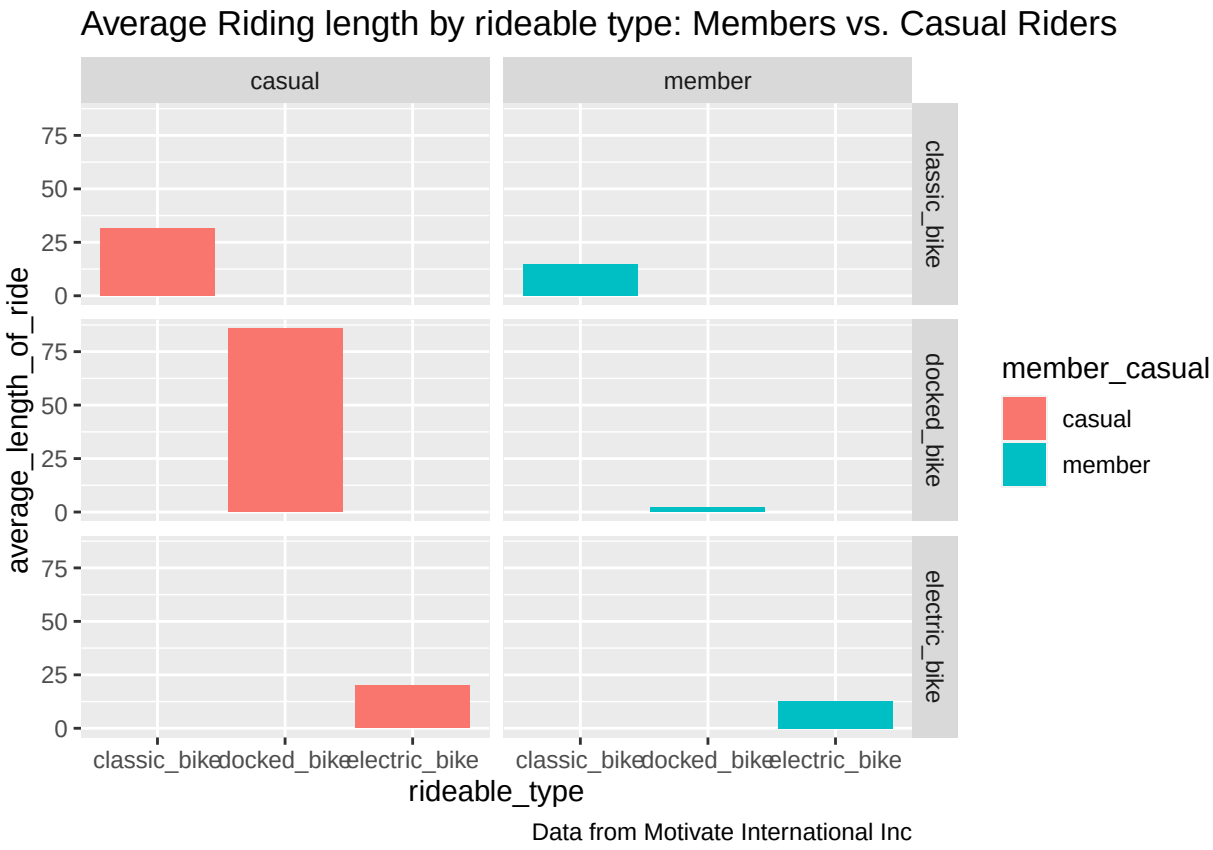
### Calculate the average length of ride by rideable type

```
Cleaned_all_bike_share_summary1<-Cleaned_all_bike_share%>%
  group_by(rideable_type, member_casual)%>%
  summarize(average_length_of_ride=mean(Length_of_ride), .groups='drop')
```

### data visualize

```
ggplot(Cleaned_all_bike_share_summary1, aes(x=rideable_type, y=average_length_of_ride, fill=member_casua
  geom_col()+
  labs(title ="Average Riding length by rideable type: Members vs. Casual Riders", caption ="Data from
  facet_grid(rideable_type~member_casual)
```

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```



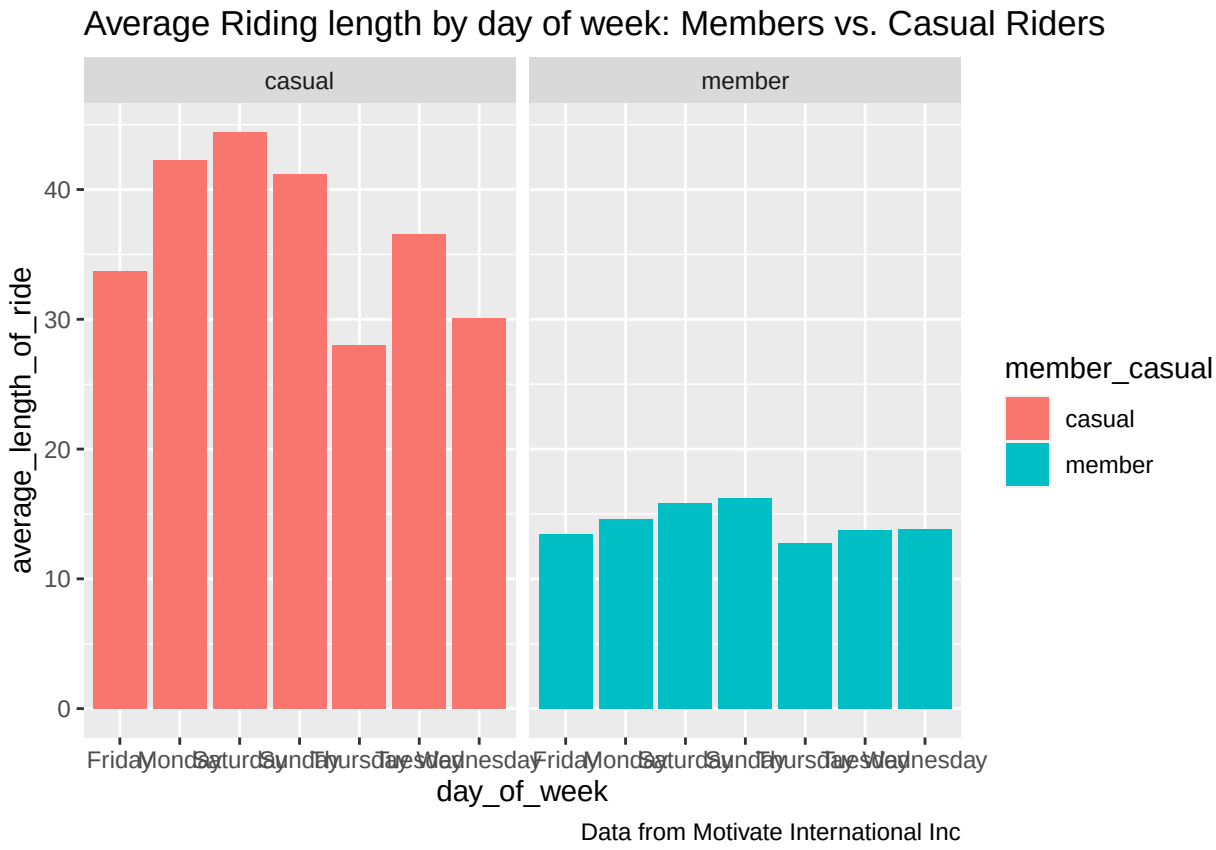

## Caculate the averge length of ride by day of week

```
Cleaned_all_bike_share_summary2<-Cleaned_all_bike_share%>%
  group_by(day_of_week, member_casual)%>%
  summarize(average_length_of_ride=mean(Length_of_ride), .groups='drop')
```

## data visualization

```
ggplot(Cleaned_all_bike_share_summary2, aes(x=day_of_week, y=average_length_of_ride, fill=member_casual
  geom_col()+
  labs(title ="Average Riding length by day of week: Members vs. Casual Riders", caption ="Data from Mo
  facet_wrap(~member_casual)
```

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```



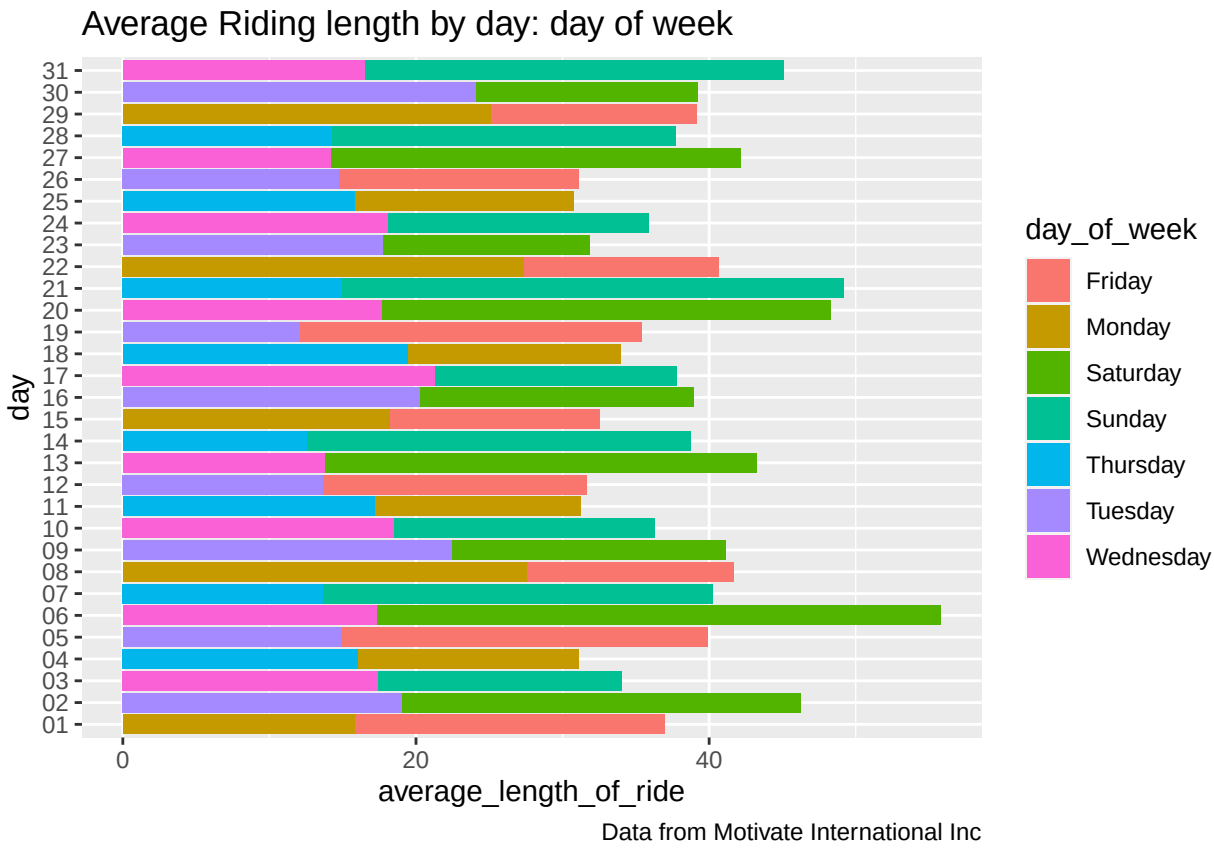

### Calculate the averag length of ride by day

```
Cleaned_all_bike_share_summary3<-Cleaned_all_bike_share%>%
  group_by(day, day_of_week)%>%
  summarize(average_length_of_ride=mean(Length_of_ride), .groups='drop')
```

### data visualization

```
ggplot(data=Cleaned_all_bike_share_summary3)+
  geom_col(mapping=aes(x=average_length_of_ride, y=day, fill=day_of_week))+
 labs(title ="Average Riding length by day: day of week", caption ="Data from Motivate International In
```

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```




## Calculate top 15 started stations by average ride of length

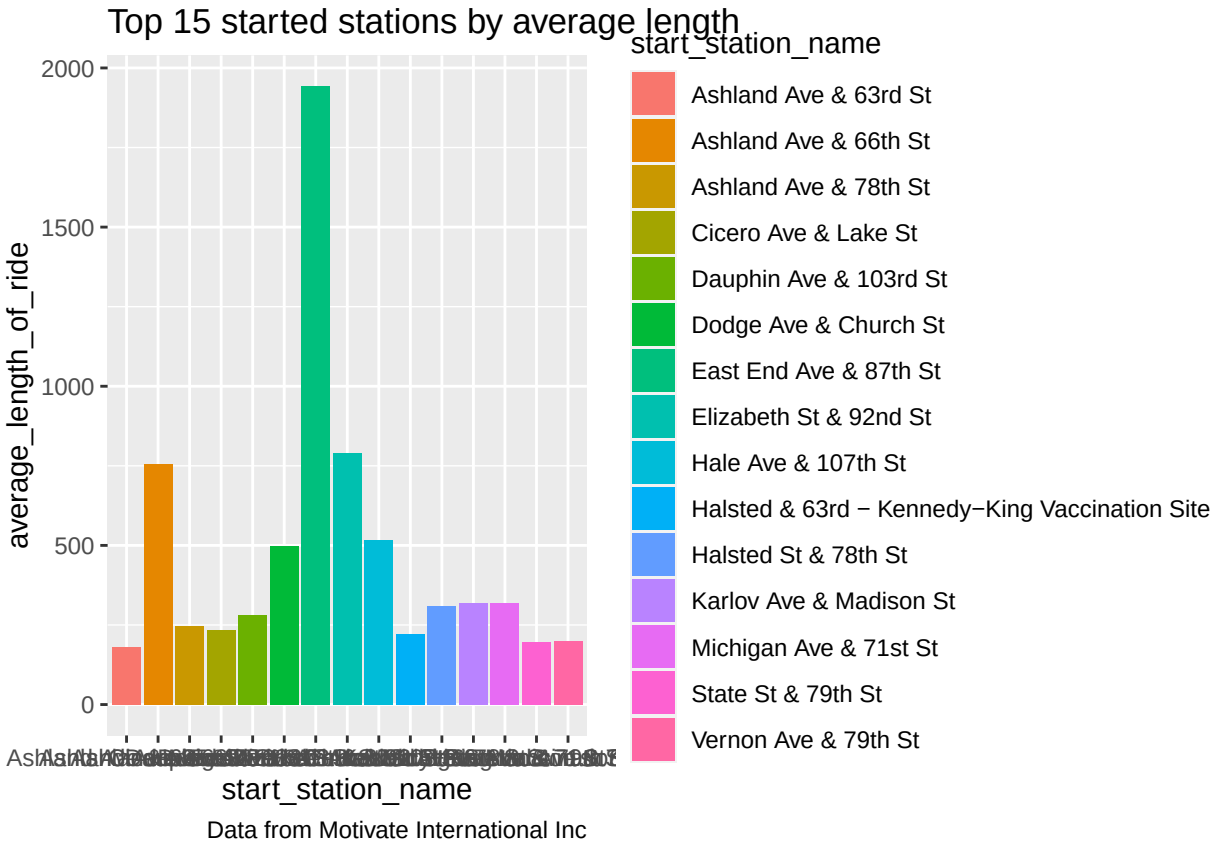
```
Cleaned_all_bike_share_summary4<-Cleaned_all_bike_share%>%
  group_by(start_station_name)%>%
  summarize(number_of_started_at=nrow(start_station_name),average_length_of_ride=mean(Length_of_ride))%
  arrange(desc(average_length_of_ride))

Cleaned_all_bike_share_summary5<-head(Cleaned_all_bike_share_summary4, n=15)
```

## data visualization

```
ggplot(data=Cleaned_all_bike_share_summary5)+
  geom_col(mapping=aes(x=start_station_name, y=average_length_of_ride, fill=start_station_name))+
 labs(title ="Top 15 started stations by average length", caption ="Data from Motivate International In
```

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```




## Key findings

**–Casual riders' average ride duration is higher than member riders.**

**–The average ride duration of Casual riders during weekends is longer than weekdays.**

**–Docked bikes are the most popular compared with other bikes.**

**–Top 15 stations by average length of ride.**

## Process 5: Share

### Recommendations

**–Put more docked bicycles at Top 15 stations to further increase the number of casual riders**

**–offer membership discount to casual riders on Saturday and Sunday to attract them transfer to anual member**

**–Provide two kinds of offer for the membership, one is anual member, the other one is according to the average length of ride since the average ride duration is longer than anual member.**

## Further analsis

**I will further analyze the riding behavior if there is data about riders.**